A Main Appendix

Contents

A.1 Data Sources

Table A.1: Data Sources

Data	Source	Access
Weather	Institute of Hydrology, Meteorology and Environmental Studies (IDEAM)	ideam.gov.co
Soil	Soil and Terrain Database for Latin America and the Caribbean (SOTERLAC)	data.isric.org
Census	National Administrative Department of Statistics (DANE)	datosabiertos.esri.co; dane.gov.co
Terrain	Multi-Error-Removed Improved Terrain DEM	hydro.iis.u-tokyo.ac.jp
Highways	National Roads Institute (INVIAS)	inviasopendata-invias.opendata.arcgis.com
Drug Cultivation and Seizures	SIDCO Database, Ministry of Justice and Law (MJD)	minjusticia.gov.co
Flooding	National Unit for Disaster Risk Management (UNGRD)	portal.gestiondelriesgo.gov.co
Crops	Ministry of Agriculture and Rural Development (MADR)	agronet.gov.co
Armed Group Presence	Institute for Peace and Development Studies (Indepaz)	indepaz.org.co
Armed Group Presence Survey	Mapping Attitudes, Perceptions and Support (MAPS) dataset	Weintraub et al. (2023)
Transit Cost Raster	Malaria Atlas Project (MAP)	malariaatlas.org
Ports	World Port Index (WPI)	msi.nga.mil
Military Bases	Colombian Army, Airforce and Navy	datos.gov.co; fac.mil.co; armada.mil.co
Election Results (2010-2018)	National Election Registry	registraduria.gov.co
Election Results (1986)	University of the Andes CEDE Database	Pachón, Sánchez Torres, et al. (2014)

A.2 Violence by Splinter Groups

Table A.2: CNMH Conflict Events Involving FARC Splinter Groups, 2017-2020

Incident Type	Count
Military Engagements with State Security Forces	115
Incidents of Sexual Violence	110
Selective Assassinations	93
Landmines and Explosives	60
Military Engagements with Non-State Actors	44
Kidnapping or Hostage-Taking	31
Attacks on Infrastructure	30
Incidents of Forcible Recruitment	14
Forced Disappearances	12
Massacres (i.e. 3 or more victims)	5

Notes: Data comprise all incidents from the CNMH Sievcac database in which FARC splinter groups are listed as the presumed perpetrator. For battles, this includes incidents where FARC splinter groups are one of the belligerents.

A.3 Chronology of Splinter Group Formation

Formed		Group	Founder	Rank	Tier	Alliance	Notes
2016	*Summer*	1st Front	Iván Mordisco	Front-level	Low	EMC	
		GUP	Don Y	Front-level	Low	None	
	Winter	Southeastern Bloc	Gentil Duarte	General Staff	Middle	EMC	
		Frente Acacio Medina	John 40	Front-level	Low	SM	*Switched alliances*
2017	*Spring*	Jaime Martinez Column	Mordisco	Front-level	Low	EMC	
	Summer	Estiven González Front	Sábalo	Front-level	Low	None	
		Jacopo Arenas Column	Pija	Front-level	Low	None	
	Fall	Oliver Sinisterra Front	Gaucho	Front-level	Low	None	
		48th Front	Sinaloa	Front-level	Low	SM	
2018	*Winter*	36th Front	Cabuyo	Front-level	Low	EMC	
		33rd Front	Jhon Milicias	Front-level	Low	EMC	
		10th Front	Jerónimo	Front-level	Low	EMC	
	Spring	18th Front	Ramiro	Front-level	Low	SM	*Split from 36th; switched alliances*
	Summer	Rafael Aguilera Front	Chumbi	Front-level	Low	EMC	
		Dagoberto Ramos Front	El Indio	Front-level	Low	EMC	
	Fall	Carlos Patiño Front	Gentil Duarte	General Staff	Middle	EMC	*Franchised to Mauricio Córdoba*
2019		28th Front	Antonio Medina	Front-level	Low	EMC	
	Spring	Carolina Ramírez Front	Gentil Duarte	General Staff	Middle	EMC	*Franchised to Alonso 45*
	Summer	Segunda Marquetalia	Iván Márquez	Secretariat	High	SM	
	Fall	Bloque Alfonso Cano	Allende	Front-level	Low	SM	*Split from Oliver Sinisterra*

Note: Major groups are groups whose maximum estimated size reached 100 guerrillas. A comprehensive list of sources can be accessed online.

A.4 Alternative Dependent Variable Specification

Table A.3: Effects of coca cultivation on FARC splinter group emergence by 2017

	Dependent variable: Splinter Group Presence by 2017 (binary)				
	(1) Difference-in-Means	(2) Difference-in-Means	(3) Covariate-Adjusted	(4) Instrumental Variables	(5) Instrumental Variables
(Intercept)	0.11^{***}	0.07^{**}	0.04	-0.02	-0.03
	(0.02)	(0.03)	(0.02)	(0.14)	(0.13)
Coca Cultivation	0.39^{***}		0.46^{***}	0.39^{***}	
Binary	(0.06)		(0.05)	(0.08)	
Coca Cultivation		0.06^{***}			0.06^{***}
Continuous		(0.01)			(0.01)
Agriculture Control				✓	✓
Flood Control				✓	✓
Num. obs.	260	260	258	260	260
First-stage F-stat				240.27	242.14
Anderson-Rubin CI				$[0.23, 0.56]$	$[0.03, 0.08]$

Notes: The dependent variable is a binary indicator for the presence of FARC splinter groups in a municipality as of 2017. The independent variable in models 1 and 3 is the logged maximum annual coca cultivation (in hectares) in a municipality between 1999-2012, and in models 2, 4, and 5 a binary indicator that takes a value of one if annual coca cultivation exceeded 100 ha. in any year during this period. The full sample consists of 260 FARC municipalities, with 2 observations dropped on model 3 due to missingness in the covariates. In model 3 I generate weights on the covariates of interest (literacy, electricity, population, percent rural, rough terrain, highway coverage, and the three distance variables) with kernel balancing, and estimate the model using weighted least squares. The instrument in models 4 and 5 is a municipality-level index of soil (acidity, nitrogen, carbon, and drainage) and weather (temperature, sunlight hours, annual rainfall, and humidity) variables combined using OLS in for the continuous coca measure, and Logit for the binary measure. The IV models include controls for agricultural productivity, defined as the logged maximum annual hectares under cultivation with legal crops between 2009 and 2014, and flooding, defined as the average number of flood-related alerts per municipality per year between 2017-2020. All models use HC2 heteroskedasticity consistent standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.5 Dependent Variable Validation

Table A.4: VALIDATING SPLINTER GROUP PRESENCE WITH SURVEY DATA

	Dependent variable: Proportion of Respondents Reporting FARC Splinter Group Control
	(1)
(Intercept)	0.17^{***}
	(0.03)
Splinter Group Presence Measure	0.17^{***}
	(0.04)
Num. obs.	54
F statistic	17.46

Notes: Model 1 is an OLS regression where the dependent variable is the proportion of survey respondents reported that FARC splinter groups were in charge in their municipality, based on data from the first wave of the MAPS survey which was enumerated in 2019 (Weintraub et al. 2023). This proportion omits "don't know" responses. Survey data were available for 54 of the municipalities included in my data. The independent variable is a binary measure of whether FARC splinter groups were present in a municipality by 2020 based on data from Indepaz. I use HC2 standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.6 Covariate Adjustment

Figure A.1: Balance on Observables for Drug Producing and non-Drug Producing Municipalities

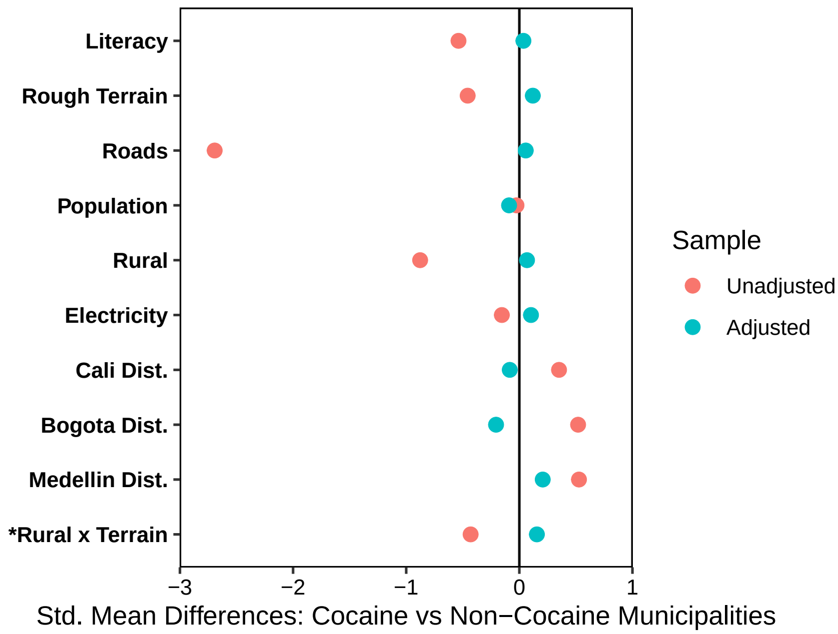
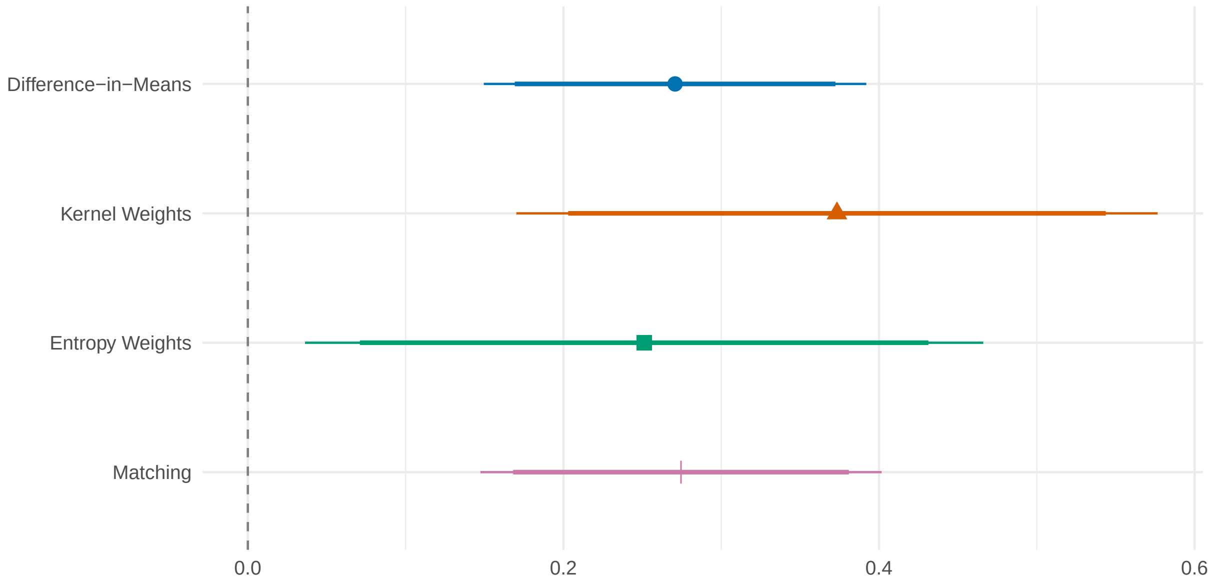


Table A.5: Balancing Weights at Varied kbal Bandwidths

Bandwidth	Estimate of Y $\sim$ D	Effective Sample Size (Control)	Omnibus F-stat for D $\sim$ X
15.37	0.37	12.13	0.78
10.00	0.39	12.37	0.37
5.00	0.42	17.76	4.86
2.50	0.44	20.06	4.97

Notes: Kernel balancing allows for manual tuning of the bandwidth scaling factor in the calculation of gaussian kernel distance equivalent to the entire denominator $2\sigma^2$ of the exponent. The default is to search for the value which maximizes the variance of the kernel matrix. Here, each row shows the results of manually varying this parameter, with the second column showing the weighted estimate for outcome of splinter group presence variable on the binary treatment of coca cultivation, the second column shows the effective sample size of the weighted control units, and the fourth column shows the omnibus F-statistic for the weighted regression of the treatment variable on all covariates.

Figure A.2: Comparing Estimates for the Effect of Coca Cultivation with Various Weighting and Matching Techniques



Notes: In all models, the dependent variable is a binary indicator for the presence of FARC splinter groups in a municipality as of 2020, and the independent variable is a binary indicator for coca cultivation, which takes a value of 1 if the municipality's annual coca cultivation ever reached at least 100 hectares, and 0 otherwise. The full sample consists of 260 FARC municipalities, with 2 observations dropped in the covariate-adjusted models due to missingness. In the Kernel balancing and Entropy Balancing models, I generate weights on the covariates of interest (literacy, electricity, log population, percent rural, rough terrain, highway coverage, and the three distance variables) with kernel balancing and entropy balancing respectively (Hazlett 2020; Hainmueller 2012), and estimate the model for each subsample using weighted least squares. In the matching model, I match on these variables using mahalanobis-distance nearest-neighbor matching. In generating the matching and balancing methods, the target estimand is the ATT, which means that weights play the role of making the control units more similar to the treated. Thick and thin error bars represent 95% and 90% confidence intervals respectively. All models used HC2 heteroskedasticity consistent standard errors.

A.7 Production Instrument Data

A.7.1 Construction

Table A.6: Coefficients for Regression-based Instrument Construction

	Dependent Variable:	
	Coca Production (Log ha.)	*Coca Producer (Binary)*
	(1)	(2)
(Intercept)	1.822	3.416
	(2.301)	(2.441)
Soil PH	-1.340^{***}	-1.803^{***}
	(0.305)	(0.373)
Temperature Range	-0.001	-0.009
	(0.027)	(0.033)
Soil Nitrogen	0.098	0.015
	(0.127)	(0.126)
Soil Carbon	-0.002	-0.001
	(0.002)	(0.002)
Sunlight Hours	0.942^{**}	0.709^{*}
	(0.301)	(0.313)
Humidity Range	-0.030^{***}	-0.050^{***}
	(0.007)	(0.015)
Annual Rainfall	0.002^{***}	0.001^{**}
	(0.000)	(0.000)
Soil Drainage	3.163^{***}	2.082^{***}
	(0.487)	(0.556)
Num. obs.	260	260

Notes: Model 1 is an OLS regression where the dependent variable is a continuous measure of coca cultivation (log max annual production pre-2012) and the independent variables are the soil and weather variables that comprise the instrument. Model 2 is a logistic regression where the dependent variable is a binary measure of coca cultivation (max annual production pre-2012 $>$ 100 in any year) and the independent variables are the soil and weather variables that comprise the instrument. Both models use HC2 standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.7.2 Output

Figure A.3: Distribution of Suitability Index (OLS)

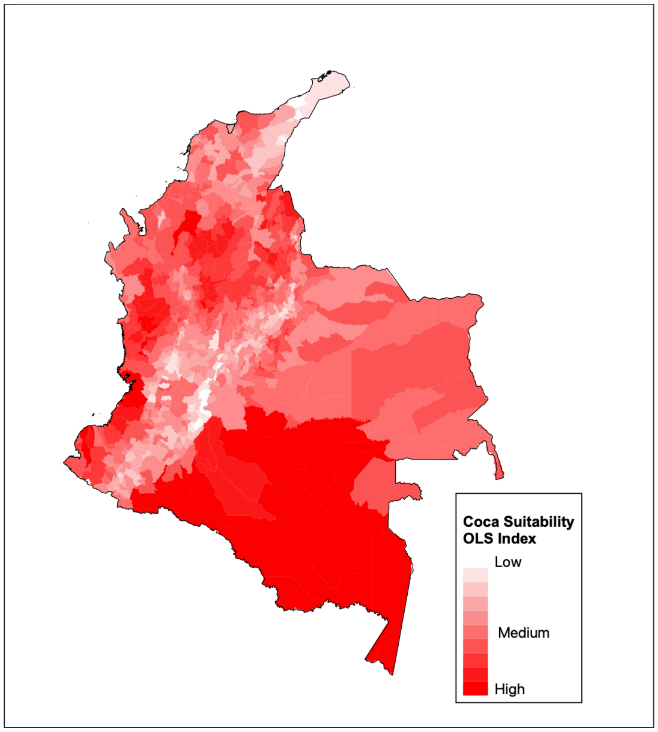
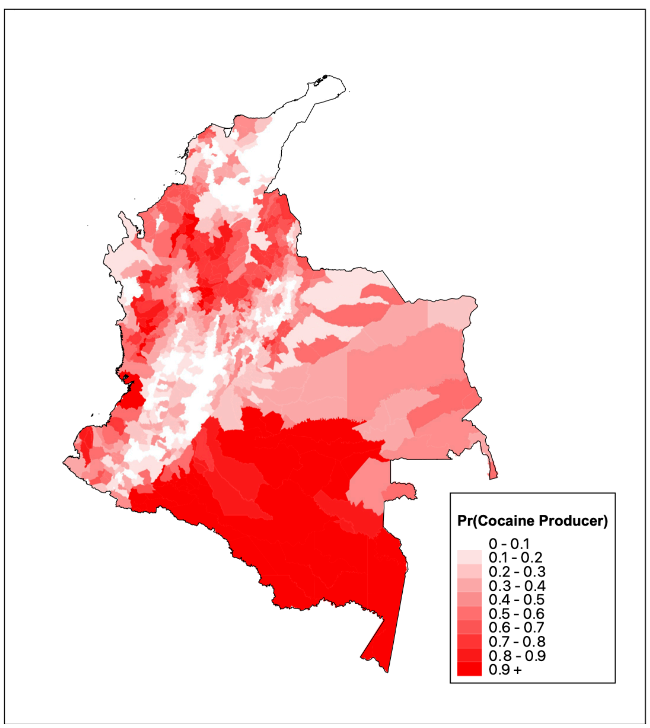


Figure A.4: Distribution of Suitability Index (Logit)



A.8 Production Robustness Checks

Table A.7: Production IV: Drop Rain from Weather & Soil Index

	(1)	(2)	(3)	(4)
(Intercept)	0.19^{***}	-0.22	0.31^{***}	-0.10
	(0.05)	(0.23)	(0.05)	(0.23)
Coca Cultivation (log ha.)	0.07^{***}	0.07^{***}		
	(0.01)	(0.01)		
Coca Cultivation (binary)			0.27^{**}	0.28^{**}
			(0.10)	(0.10)
Agriculture Control		0.05		0.05^{*}
		(0.03)		(0.03)
Floods Control		-0.03		-0.05
		(0.07)		(0.07)
Num. obs.	260	260	260	260
First-stage F-stat	169.33	169.77	234.90	231.02

Notes: The dependent variable is a binary indicator for the presence of FARC splinter groups in a municipality as of 2020. The independent variable is a binary indicator that takes a value of one if annual coca cultivation in a municipality exceeded 100 hectares between 1999-2012. The instrument is a municipality-level index of soil (acidity, nitrogen, carbon, and drainage) and weather (temperature, sunlight hours, and humidity) variables combined using predictions from ols or logistic regression. The 2SLS model with controls includes controls for agricultural productivity, defined as the logged maximum annual hectares under cultivation with legal crops between 2009 and 2014, and flooding, defined as the average number of flood-related alerts per municipality per year between 2017-2020. All models use HC2 heteroskedasticity consistent standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.8: IV Reduced Form with Spatial (Conley) Standard Errors

	Dependent Variable: *Splinter Group Presence (Binary)*				
	(1)	(2)	(3)	(4)	(5)
Radius	1000km	750km	500km	200km	100km
(Intercept)	0.25***	0.25***	0.25**	0.25**	0.25**
	(0.06)	(0.07)	(0.08)	(0.09)	(0.08)
Weather & Soil Instrument	0.05*	0.05*	0.05	0.05*	0.05**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Num. obs.	260	260	260	260	260

Notes: The instrument is a municipality-level index of soil (acidity, nitrogen, carbon, and drainage) and weather (temperature, sunlight hours, annual rainfall, and humidity) variables combined using principal components analysis. The reduced form model regresses splinter group presence, a binary indicator for the presence of FARC splinter groups in a municipality as of 2020, on the instrument. All models use Conley standard errors to adjust for potential spatial dependence in the error term (Conley 1999). I use municipality centroids and distance cutoffs for potential clustering of 100, 200, 500, 750, or 1000 km. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.9 Sensitivity Analysis

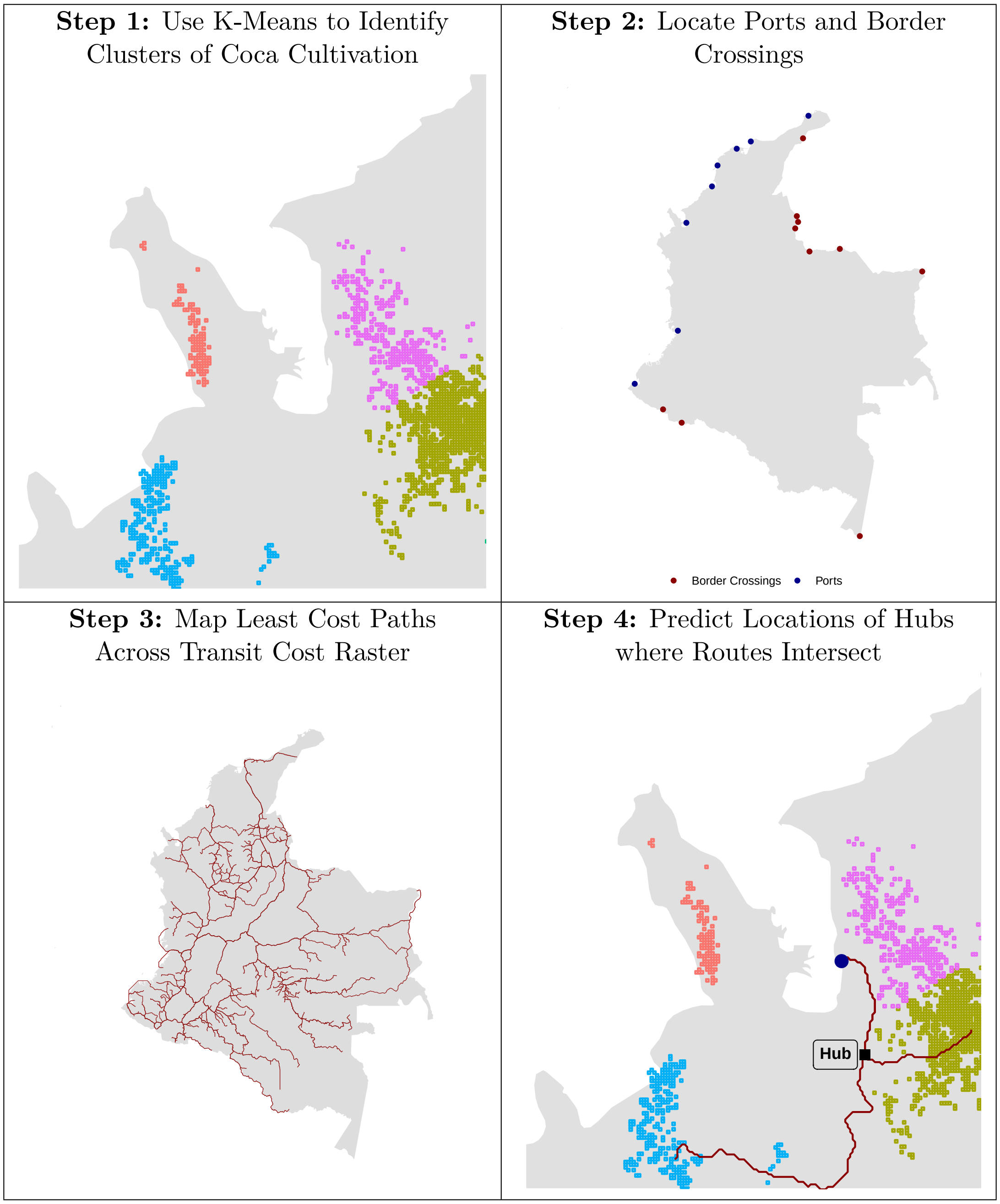
Table A.9: Sensitivity Analysis for Reduced Form of IV for Coca Cultivation

	Outcome: *FARC Splinter Group Presence (binary)*						
	Est.	S.E.	t-value	$R^2_{Y\sim D\|\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$	DF
Weather & Soil Instrument	0.05	0.01	4.18	6.3%	22.8%	12.8%	258

Notes: The dependent variable is a binary indicator for the presence of FARC splinter groups in a municipality as of 2020. The instrument is a municipality-level index of soil (acidity, nitrogen, carbon, and drainage) and weather (temperature, sunlight hours, annual rainfall, and humidity) variables, combined using the predicted values from an OLS regression on historical coca cultivation. I use HC2 heteroskedasticity consistent standard errors.

A.10 Trafficking Instrument Data

Figure A.5: Measurement Strategy



Notes: For Step 1, I aggregate remote sensed coca cultivation sites between 2001 and 2015, and use k-means clustering to identify 200 clusters. I use the cluster centroids as the starting points. For Step 2, I identified Colombian ports using the World Port Index, and border crossings using points where the Colombian highway network came within 2 miles of the border. Where multiple ports or border crossings overlap, I randomly choose one unique point. For step 3, I use the Dijkstra's algorithm to map to route from every coca cultivation cluster to every border crossing and port across a raster of transit costs that has unique travel speeds for roads, rivers, mountains, and other terrain features. I keep the least cost path from each cluster to one port and one border crossing. Step 4 locates all points where two or more routes intersect to generate a count of the number of hubs per municipality.

A.11 Alternative Explanations

A.11.1 Alternative Measure for Terrain Ruggedness

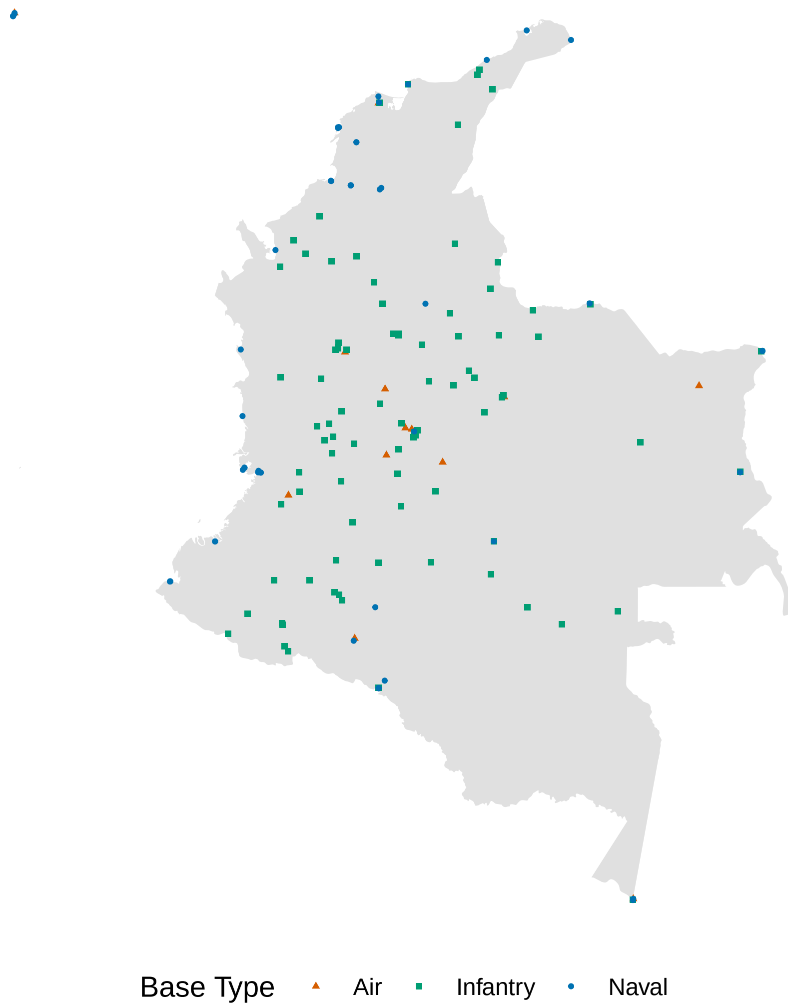
Table A.10: Alternative Measure of Terrain Ruggedness

	(1) OLS	(2) OLS	(3) IV
(Intercept)	0.42***	0.42***	0.42***
	(0.03)	(0.03)	(0.03)
Terrain Ruggedness (TRI)	−0.03	0.02	0.01
	(0.03)	(0.03)	(0.03)
coca cultivation		0.18***	0.13***
		(0.04)	(0.03)
Num. obs.	260	260	260
F statistic	1.10	9.80	9.69

Notes: The dependent variable is a binary indicator for the presence of FARC splinter groups in a municipality as of 2020. In the IV model, the instrument is a municipality-level index of soil (acidity, nitrogen, carbon, and drainage) and weather (temperature, sunlight hours, annual rainfall, and humidity) variables, combined using the predicted values from an OLS regression on historical coca cultivation. The variable for coca cultivation is the logged maximum hectares of coca cultivation between 1999-2012. The TRI measure of terrain ruggedness applies the Riley, DeGloria, and Elliot (1999) index of topographical heterogenity to a high-resolution global digital elevation model (90m at the equator) developed by Yamazaki et al. 2017, which is then aggregated to produce a ruggedness score for each municipality. All independent variables are scaled. All models use HC2 heteroskedasticity consistent standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.11.2 Distribution of Colombian Military Bases

Figure A.6: Map of Colombian Military Bases



Notes: These data are derived from the Colombian Army's list of Army Base Infirmaries (Establecimientos de Sanidad Militar, Ejército Nacional), the Colombian Airforce's georreferenced list of Air Force Units (Unidades Aéreas), and the Colombian Navy's list of Naval Units (Unidades Armada Nacional). Note that some Colombian naval bases are located along major rivers.

A.11.3 Alternative Political Support Measure

Table A.11: FARC Political Wing 1986 Voteshare & Splinter Group Emergence

	Dependent variable: Splinter Group Presence (binary)		
(Intercept)	0.42***	0.42***	0.42***
	(0.03)	(0.03)	(0.03)
FARC Political Wing Voteshare	0.13***	0.10***	0.08**
	(0.02)	(0.03)	(0.03)
Coca Cultivation		0.10**	0.14***
		(0.03)	(0.04)
Num. obs.	260	260	260
F statistic	28.06	21.72	22.31

Notes: The dependent variable is a binary indicator for the presence of FARC splinter groups in a municipality as of 2020. In the IV models, the instrument is a municipality-level index of soil (acidity, nitrogen, carbon, and drainage) and weather (temperature, sunlight hours, annual rainfall, and humidity) variables, combined using the predicted values from an OLS regression on historical coca cultivation. The variable for coca cultivation is the logged maximum hectares of coca cultivation between 1999-2012. The measure of support for the FARC's political wing in all three models is the municipality-level voteshare for Jaime Pardo Leal, the candidate representing the FARC's political wing—Unión Patriotica—in Colombia's 1986 presidential election. For a small number of municipalities created between 1986 and 2018, I use the voteshare from the municipality(s) from which they were split. All models use HC2 heteroskedasticity consistent standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.12: Support for FARC Political Wing in 1986 & Contemporary Voteshare of Left Candidates

	DV = Left Candidate Voteshare: 2010	DV = Left Candidate Voteshare: 2014	DV = Left Candidate Voteshare: 2018
	Model 1	Model 2	Model 3
(Intercept)	0.20***	0.48***	0.35***
	(0.00)	(0.01)	(0.01)
FARC Political Wing	0.14***	0.19***	0.23***
Voteshare (1986)	(0.04)	(0.04)	(0.04)
Num. obs.	1112	1113	1113
F statistic	12.26	21.45	30.06

Notes: The independent variable in all three models is the municipality-level voteshare for Jaime Pardo Leal, the candidate representing the FARC's political wing—Unión Patriotica—in Colombia's 1986 presidential election. The dependent variable in each model is the municipality-level two-party voteshare for the presidential candidate further to the left in the second round of the 2010, 2014, and 2016 elections respectively. This corresponds to Antonius Mockus in 2010, Juan Manuel Santos in 2014, and Gustavo Petro in 2018. For a small number of municipalities created between 1986 and 2018, I use the voteshare from the municipality(s) from which they were split. All models use HC2 heteroskedasticity consistent standard errors. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Appendix References

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45.

Hainmueller, Jens. 2012. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis* 20 (1): 25–46.

Hazlett, Chad. 2020. "Kernel balancing." *Statistica Sinica* 30 (3): 1155–1189.

Pachón, Mónica, Fabio José Sánchez Torres, et al. 2014. "Base de datos sobre resultados electorales CEDE, 1958-2011."

Riley, Shawn J, Stephen D DeGloria, and Robert Elliot. 1999. "Index that quantifies topographic heterogeneity." *Intermountain Journal of Sciences* 5 (1-4): 23–27.

Weintraub, Michael, Abbey Steele, Sebastián Pantoja-Barrios, Håvard Mokleiv Nygård, Marianne Dahl, and Helga Malmin Binningsbø. 2023. "Introducing the Mapping Attitudes, Perceptions and Support (MAPS) dataset on the Colombian peace process." *Journal of Peace Research,* 00223433231178848.

Yamazaki, Dai, Daiki Ikeshima, Ryunosuke Tawatari, Tomohiro Yamaguchi, Fiachra O'Loughlin, Jeffery C Neal, Christopher C Sampson, Shinjiro Kanae, and Paul D Bates. 2017. "A high-accuracy map of global terrain elevations." *Geophysical Research Letters* 44 (11): 5844–5853.